

## **Development and validation of a pain behavior assessment in patients with chronic low back pain**

Katharina Meyer<sup>1)</sup> , Andreas Klipstein<sup>2)</sup> , Peter Oesch<sup>3)</sup> , Beatrice Jansen<sup>4)</sup> , Jan Kool<sup>3,5)</sup> , Karin Niedermann<sup>5)</sup>

1) Physiotherapy & Occupational Therapy, University Hospital Zurich, Zurich, Switzerland

2) Department of Rheumatology University Hospital Zurich, Zurich, Switzerland and Center of Occupational Health, Militärstrasse 76, 8004 Zurich, Switzerland  
klipstein@aeht.ch, MD MSc

3) Research Department and Department of Rheumatology, Rehabilitation Centre Valens, 7317 Valens, Switzerland  
peter.oesch@kliniken-valens.ch, PhD, jan.kool@kliniken-valens.ch, PhD

4) Department of Work Rehabilitation, Rehaklinik Bellikon, suva care, 5454 Bellikon, Switzerland  
beatrice.jansen@rehabellikon.ch, MSc

5) Zurich University of Applied Sciences, School of Health Professions, Institute of Physiotherapy, 8400 Winterthur, Switzerland  
nika@zhaw.ch, PhD

### **Corresponding author:**

Katharina Meyer

University Hospital Zurich, U OST 153, Gloriastr. 25, 8091 Zurich, Switzerland

E-mail: Katharina.Meyer@usz.ch, Phone: +41 44 255 36 17, Fax: +41 44 255 43 88

**Key words:** Chronic pain; low back pain; pain measurement; reproducibility of results; return to work

## Introduction

Patients with chronic low back pain (LBP) suffer from physical and psychological distress and disability. Multidisciplinary rehabilitation and functional rehabilitation programs (FRP) [1] for such patients with physically demanding jobs focus on musculoskeletal reconditioning, work reintegration and behavioral training, but are time-consuming and expensive [2, 3]. Patients with physically demanding jobs are often migrants with lower socio-economic status, different cultural backgrounds and language barriers. Thus, psychosocial constraints may influence both the rehabilitation process and return to work adversely [4-6]. Psychosocial stressors in patients with chronic non-specific low back pain (CNSLBP) may be accompanied by high levels of pain behavior, that may not reflect the limitations based on organic-structural changes [7, 8]. Single signs of pain behavior are commonly present in patients with CNSLBP, but a high number of pain behavior signs influences return to work adversely [9]. Previous research has used different terms such as exaggerated pain behavior, sincerity of effort and symptom magnification. These terms are criticized for suggesting intention of patients [10]. Therefore, we decided to use the more neutral term “pain behavior” and, in the case of the presence of many signs, “high levels of pain behavior”. Pain behavior should be considered in all patients during the evaluation of functional performance and, furthermore, in the determination of disability allowances[11]. Prior to FRP, an efficient screening process should identify patients with low pain behavior who may benefit from FRP and, thereby, prevent the inclusion of patients who are unlikely to benefit from such expensive programs. In patients with high levels of pain behavior, interventions with focus on coping and cognitive behavioral treatment should be applied [12, 13]. Thus, assessments are needed to effectively allocate the right patients to the right rehabilitation programs.

Several authors have recommended combining assessments [9, 14, 15] of pain behavior and effort [16] during physical performance tests in order to increase the predictive power in the selection process of patients for FRP [17-19]. Examples of such assessments are the *Spinal Function Sort* (SFS) [20], grip strength [14] or the *nonorganic somatic components* by Waddell [17, 21, 22].

A comprehensive assessment to screen for high levels of pain behavior should include different aspects of pain behavior [14] such as: a) *pain perception and description by the patient*; b) *the behavior of the patient* perceiving pain; c) *the patient's effort to perform physically demanding tests*; and d) *the patient's consistency of behavior* [23]. However, validated assessments covering all these different aspects do not exist to date [24]. Therefore, the pain behavior assessment (PBA) was developed to screen for patients with predominant pain behavior and help in the allocation of patients to the most effective type of therapy. The PBA is conducted principally in relation to *Functional Capacity Evaluation* (FCE) [25, 26]. The FCE is a standardized battery of clinical tests to measure a patient's physical ability to carry out work-related activities safely and, additionally, is often used to evaluate allowances for disability claims or as a measurement prior to FRP. Furthermore, the FCE contains an *assessment of inconsistency*, which has not been validated as yet. The assessment of inconsistency observes whether the patient's performance during similar physical tasks are consistent with each other [27], e.g. inability to tolerate low back flexion during the patient's examination would be reproduced during the FCE subtest through a forward bending exercise. The presence of three or more inconsistencies may indicate lack of consistency.

The aim of this study was to establish the construct validity of a German version of the PBA. The hypothesis was that the PBA is a unidimensional construct and,

therefore, a valid measurement to assess pain behavior during the evaluation of patients for FRP.

## Methods

### Participants

#### Experts:

Six experts from the association *Community of Interest for Ergonomics* (*Verein Interessengemeinschaft Ergonomie*), with at least 5 years work experience in the field of FCE, contributed to the development of the preliminary PBA, primarily in regard to the clinical relevance of the suggested items. The experts were clinicians in the field of work rehabilitation and included a physician specialized in rehabilitation and insurance medicine. This physician was an instructor of FCE and experienced in determining work disability and allowances for the Swiss Accident Insurance Fund (Schweizerische Unfallversicherung SUVA).

#### Patients:

The study was conducted in two rehabilitation clinics and in one University Hospital outpatient clinic in Switzerland. Patients with chronic non-specific low back pain (CNSLBP) [28] and with a referral to undergo FCE for the determination of fitness-for-work were included consecutively by applying the inclusion- and exclusion criteria in a stringent way. Chronic pain is defined as pain lasting for more than 3 months [29]. Further inclusion criteria were: age from 20 to 60 years and oral comprehension of German, Italian or French. Exclusion criteria were: specific LBP, by reason of nerve root compression, structural instability of the spine, tumor, inflammatory diseases, infection, spondylolisthesis, spinal stenosis, other comorbidity affecting work ability (psychiatric, cardiorespiratory, or other musculoskeletal diseases) [28].

Assessors:

The assessors were physiotherapists, who rated the PBA items through observation of the pain behavior of patients during FCE-testing. All FCE assessors had completed a standardized two-day course on FCE and were qualified FCE-testers, with 2-15 years of experience in FCE testing. In addition, assessors were trained during a 4-hour session at all centers in the application and interpretation of the PBA items, including a discussion of 8 FCE video sequences, by the same FCE instructor. On this basis, disagreement between assessors was detected and solved. The training protocol contained interpretation guidelines for difficult items and the protocol was based on the manual. It has been shown in a recent study that the inter-rater reliability for effort determination of FCE tests is on average 0.73 [30].

## **Study design**

A cross-sectional study design was performed.

### **Steps:**

Figure 1 shows the consecutive steps in the development and validation process of the PBA.

#### *Step 1) Development of the preliminary PBA:*

The PBA was mainly developed by an expert group on the basis of the *assessment of inconsistency* [23], as a part of the FCE. The findings of Keefe [8] provided 6 further items and the *nonorganic signs* of Waddell [21] eight items. Another 28 items relevant to pain behavior were added from various sources [14,23,27,31]. The experts added a further ten items, considered relevant, from their own clinical experience and suggested the following four subscales to differentiate between the important aspects of pain behavior, as observed in clinical practice: a) *pain perception*; b) *overt pain behavior*; c) *effort*; and d) *consistency of behavior*.

A dichotomous scale (0=no pain behavior present; 1= pain behavior present) was applied to all items. The criteria for rating pain behavior “yes/no” were defined by numeric cutoff levels (where applicable), evaluated in previous studies [15, 19, 20, 32-35]. Sensitivity and specificity values for these cutoff levels were between 0.55 to 1.0 and 0.87 to 0.97, respectively. For 3 items (*grip strength, lifting and one-hand carry*), where no published cutoffs were available, the 10<sup>th</sup> percentile was chosen as a very conservative estimate.

#### *Step 2) Pretesting:*

The preliminary PBA testing was performed in the daily clinical settings of the participating work rehabilitation centers. The feedback of the assessing physiotherapists was evaluated by the experts and guided further adaptations.

#### *Step 3) Construct validity:*

A cross-sectional study was performed in the three work rehabilitation centers of Bellikon, Valens and Zurich University Hospital in Switzerland. Eight physiotherapists administered the Functional Capacity Evaluation, including the revised PBA, to 200 patients. The study was approved by the local ethics committee and performed in accordance with the ethical standards of the Declaration of Helsinki. All patients gave their written informed consent prior to their inclusion in the study.

#### **Statistics:**

Rasch analysis was performed to test the unidimensionality of the PBA subscales, i.e. if indeed one single underlying trait in each subscale is measured [36, 37]. Rasch modelling reverses the traditional data – model relationship, i.e. providing various error estimates and fit statistics, in order to examine how well the observed data fit the Rasch model's expectations. Rasch model theory states that response categories change as a function of participant ability and item difficulty (both expressed as logits), i.e. the probability that a person with a logit of 1.0 will pass an item with a

difficulty of 1.0 is 50%, but the probability of an item with a difficulty of >1 logit or <1 logit will be <50% and >50%, respectively. Thus, each item and person is calibrated on an abstract continuum of the trait from less to more, providing an equal interval scale representing the variable, in this case pain behavior. Low ability on this scale means that no, or few, pain behavior is present; whilst a high ability means that much pain behavior is present. Using this calibration, a linear transformation of the scale is obtained, allowing for valid parametric statistics. For construct validity (step 3), data from the revised PBA were fitted to the Rasch Dichotomous Model ( $\ln(p/(1-p)) = \text{logit}$ ). Overall summary fit statistics provided *item-person interaction statistics* with mean item fit residuals standard deviations (SD), expected to be close to one given adequate fit to the model, and *chi squared item trait interaction statistics*, expected to be non-significant to demonstrate that no invariance of item difficulty across the scale is present. Item fit statistics provided information on the fit of the individual items. Items showing significant p-values (i.e. misfit = significant deviation from the observed data to the expected data) and/or fit residuals > +/- 2.5 (meaning redundancy and another dimension, respectively) were deleted [38]. Bonferroni corrections on Alpha levels set at 0.05 resulted in significance levels of 0.001. Differential item functioning (DIF) analysis was applied to assess whether uniform DIF (consistent ability differences between groups, e.g. gender) and / or non-uniform DIF (inconsistent ability differences between groups) was present. Items displaying DIF at the defined significance level are usually removed in the process of developing a new scale. DIF assessed bias for *age*, *gender*, *native language*, work ability (measured with *Work Ability Index WAI*), *fear of losing job*, *fear avoidance beliefs* (measured with *Fear Avoidance Beliefs Questionnaire FABQ*) and *duration of pain*. The values of the WAI, the FABQ, duration of pain and age were grouped in high and low levels, according to the literature [18, 39-43].

For DIF analyses, missing-values of the person factors, were substituted with the mean or, where appropriate, with the median of the sample. Missing-values of the items of the PBA were not replaced.

The internal consistency is provided by the person separation index (PSI) corresponding to Cronbach's alpha. This value should be  $\geq 0.7$  for group comparisons and  $\geq 0.85$  for individual use [36].

Principal component analysis (PCA) on the residuals from the subsets of positive versus negative loading items was applied to further test the unidimensionality of the final PBA. If this formal test is not significant, then this acts as a further proof that a unidimensional item set has been obtained.

Spearman's correlations between the final PBA total score (i.e. sum of all item scores), assuming a unidimensional version, and the Work Ability Index (WAI), the Oswestry Disability Index (ODI) and the Fear Avoidance Beliefs Questionnaire (activity scale (FABQa) and work scale (FABQw)) were calculated.

Rasch analysis was performed using the RUMM2030 software package (RUMM Laboratory, Dungraig, Western Australia). Statistical testing and correlation calculations were performed using SPSS software, version 19.0 (SPSS, Chicago, IL).

## Results

### *Step 1: Development of the preliminary PBA:*

The 10<sup>th</sup> percentile was chosen for the cutoff value for the following items: 1) *Grip strength*; 2) *Lifting*; 3) *One hand carry*. 1) *Grip strength*: In an unpublished study by Kool et al. 2005, the 10th percentile for the item was found to be 10 kg for women. Thus, scores lower than the 10<sup>th</sup> percentile was defined as high level of pain behavior; 2) *Lifting*: the cutoff was set at the 10<sup>th</sup> percentile of the lifting-weight/body-weight ratio, with values below 8% indicating the presence of pain behavior. 3) *One-*



hand carry: the 10<sup>th</sup> percentile of our own sample was chosen. Furthermore, the cutoff for the item *gait speed* was set at the 5<sup>th</sup> percentile, on the basis that it did not fit the Rasch model using the 10<sup>th</sup> percentile. However, our experts valued this item as very important clinically and, for this reason, it was kept it in the scale. The item fitted the model well using the 5<sup>th</sup> percentile (Appendix 1).

*Step 2: Pretesting:*

The preliminary PBA was applied by eight physiotherapists during clinical FCEs in 2007 and 2008. Therapists' comments resulted in an adapted 52-items version, including four new items: *denial*, *discrepancy balance test*, *discrepancy overhead work* and *coefficient of variance*. The item *Spinal Function Sort* (SFS) was adapted according to a recent study on the SFS [20]. All items generated are presented in Appendix 1. The subscale *pain perception* consisted of five items, the subscale *overt pain behavior* of seven items, the subscale *effort* of nine items and the subscale *consistency of behavior* of 31 items.

*Step 3: Construct validity:*

318 patients with LBP were referred for a fitness for work evaluation, including the FCE. In total, 118 patients were excluded based on the inclusion/exclusion criteria. 40 patients were excluded due to specific LBP, 28 due to relevant comorbidity affecting work ability, 11 due to language problems. In addition, 21 subjects did not give informed consent and 4 were older than 60 years. Moreover, 14 patients were excluded due to administrative problems. This resulted in a total of 200 patients, participating in the PBA construct validity study. Patient characteristics are shown in Table 1.

Missing values:

The proportion of missing-values to all the variables assessed for DIF was 1.7%.

Rasch analysis was performed on 51 of the 52 original items; the item (*callus despite protection of the foot*) was excluded by the RUMM software as an extreme item (never rated positive).

Firstly, data fit for each subscale was examined. Items presenting misfit and/or fit residuals  $> \pm 2.5$  (meaning redundancy and another dimension respectively) were removed, after careful consideration of its clinical relevance. The subscale *pain perception* (consisting of 5 items) fitted the Rasch model after removal of the misfit item 3 (*undifferentiated pain*): mean item fit residual was 1.35 (SD 1.04). The SD value close to 1 confirmed adequate fit to the model and chi squared item trait interaction was 14.41 (df = 8) with  $p=0.07$ , confirming no invariance of item difficulty across the scale. The subscale *overt pain behavior* (7 items) fitted the Rasch model after deletion of the misfit item 11 (*telling about pain*), resulting in a model with mean item fit residual of 0.56 (SD 0.89) and an item trait interaction of 10.03 (df = 10) with  $p=0.61$ . The subscale *effort* (9 items) fitted the model after removal of the slightly misfit item 17 (*performance of involved body part*): mean fit residual -0.16 (SD 0.75), with an item trait interaction of 14.40 (df = 16) with  $p=0.57$ . The subscale *consistency of behavior* (30 items) fitted the Rasch model after the deletion of 7 items (*further discrepancies, one hand carry, bell shape curve, rapid exchange grip, lifting floor to waist, total of nonorganic signs, general incapacity*) in a step-wise procedure. Mean fit residual was -0.36 (SD 0.89), with an item trait interaction of 51.61 (df = 46) with  $p=0.26$ .

Initial examination of the data fit of the 51 items of the PBA showed a significant item–trait interaction total chi-square of 198.80 (df = 102) with  $p<0.000$ , suggesting that there was some degree of misfit between the data and the model. This can be caused by misfit to model expectations of respondents or items, or both. The residual mean value for items was -0.43 with a SD of 1.51, an SD much higher than the

expected value of 1, i.e. no fit to the model. The high value of the SD reflects the two slightly not fitting items of the *nonorganic signs* that were left in the scale to keep the complete set of the eight *nonorganic signs* as this is an established assessment [21]. Deletion of all items, as in the analysis of each subscale, resulted in the final PBA of 41 items. The model fit had a mean item fit residual of -0.33 (SD 1.06). The chi squared item trait interaction of 100.39 (df = 82) with  $p=0.08$  confirmed that no invariance of item difficulty across the scale was present.

Individual item fit statistics ranged from  $p=0.93$  (best fit: item 39) to  $p= 0.0019$  (item 46); however, the latter still fitted the model, given the Bonferroni correction.

Individual item difficulty ranged from the lowest -3.16 (item 29), i.e. this item is scored 'easily', also with 'low ability', i.e. when no/few pain behavior is present to 2.52 (item 51) (i.e. this item is scored only with high ability, i.e. when high levels of pain behavior are present). The PSI, which corresponds to Cronbach's alpha, was 0.83. DIF analysis for all person factors revealed no bias.

Principle Component analysis (PCA) on the residuals of the positively versus the negatively loading items was slightly significant ( $p= 0.03$ ), indicating that there was some multi-dimensionality in the set of items. This was caused by the two items of the *nonorganic signs* in the subscale *consistent behavior*. When removing these two items (data not reported), the unidimensionality of the PBA was confirmed (not significant PCA,  $p=0.06$ ).

The person-item threshold distribution (Figure 2) shows that the items' difficulty (mean 0.00 (SD 1.26) did not target all persons' ability (mean -2.29 (SD 1.66)) well. The persons' ability is shifted to the left compared to the items' difficulty. This indicates that persons with low ability (i.e. low levels of pain behavior) were not targeted by the scale as there were no items capturing (very low) pain behavior. At the other end of the scale there are still items that were not scored by participants.

The median of the average total score of the final PBA was 6.5 points (IQR 2-13); the 95<sup>th</sup> percentile was 22.7 points.

The Spearman's correlation coefficients rho between the PBA and the other questionnaires were all significant with values of -0.59 for the WAI, 0.39 for the FABQa, 0.37 for the FABQw and 0.70 for the ODI.

## **Discussion**

We developed a preliminary version of the PBA and assessed 200 patients with chronic non-specific LBP. Based on RASCH analysis, considered state-of-the-art to examine construct validity of an assessment, eleven items from the initial PBA version were excluded. The final 41-item version of the PBA is a valid assessment tool to evaluate pain behavior in patients with CNSLBP.

### *Newsworthiness:*

The PBA fulfills the clinical need to comprehensively assess pain behavior, which is known to be highly predictive for the outcome of FRP. The comprehensive PBA consists of a range of aspects that so far has not been captured in one single test. The PBA considers the complexity of pain behavior and helps to screen patients for inclusion in a physically demanding FRP. Within the PBA, the items of the *assessment of inconsistency* in the FCE can be used for decisions on work disability.

### *Internal consistency:*

Items were excluded during the Rasch analysis due to statistical misfit or redundancy, but always with consideration of their clinical relevance. Therefore, two slightly misfitting items of the subscale *consistency of behavior* were kept, in favour of clinical relevance of a complete and well-established measure, but on account of confirmed unidimensionality. Thus, the decision to delete an item was not only guided

by statistical rules, but to some extent also by the opinion of FRP experts. This may increase the clinical relevance and acceptance of the PBA.

Only one item from each of the three subscales *pain perception*, *overt pain behavior* and *effort*, consisting initially of between five to nine items, were deleted during the analysis. In the largest subscale *consistency of behavior*, seven of 30 items were deleted. Three items on grip strength were excluded. This would agree with other studies that challenge the validity of grip strength measures to assess pain behavior and effort [14].

The results of the RASCH analysis support the construct of the four subscales [with some restriction in the subscale *consistency of behavior*], as well as of the whole scale. This justifies the use of each subscale and the whole scale with a total score.

The total score may be valuable for pain behavior screening in patients with CNSLBP. The scoring of the subscales may be helpful in making clinical decisions. Treatment may be specifically targeted at the major deficits identified by the corresponding subscales, e.g. a patient with a high score in the subscale *consistency of behavior* may benefit from cognitive behavioral treatment, whilst a patient with a high score in the subscale *description of pain* may need to improve coping with pain and learn about the negative impact of catastrophizing thoughts.

#### *Convergent validity:*

To evaluate the convergent validity, correlations between pain behavior and other constructs were calculated. We found moderate to good correlations between pain behavior (PBA) and the self-rated work ability (WAI) and self-rated disability (ODI). The correlation with the WAI was negative, as a high score of the WAI implies good perceived subjective work ability. We would like to point out that correlations with *fear avoidance beliefs* (FABQa and FABQw) were only moderate. This confirms that *fear avoidance beliefs* form only one aspect of pain behavior and that the construct pain

behavior is more comprehensive, including other issues such as *pain perception*, *overt pain behavior*, *effort* and *consistency of behavior*.

*Strengths of this study:*

- A large patient sample from three centers for inpatient- and outpatient functional rehabilitation was used. This increases the generalizability of the results for other patients with CNSLBP admitted for fitness for work evaluation.
- By using RASCH analysis, state-of-the-art methodology was applied when developing the new scale. This approach has more stringent criteria than factor analysis, in terms of the selection of the items [44].
- The person-item threshold distribution on the latent trait *pain behavior* was moderately targeted, with the items' difficulty mean of zero and the persons' ability mean of around -1.8. Therefore participants with no or low pain behavior may not be captured by the PBA and the PBA may neither differentiate between little and no pain behavior. However, this floor effect is not harmful to the use of the PBA, since the purpose of the PBA is to detect people with high levels of pain behavior, which is highly relevant for rehabilitation management and outcome prediction. Thus, the limited scale targeting may even help to prevent the overestimation of an individual's pain behavior.
- The PBA provides a comprehensive assessment of pain behavior. The PBA could be employed to refer patients with low levels of pain behavior and a positive prognosis for outcome to a physically demanding rehabilitation program. Alternatively, patients with higher levels of pain behavior could be referred to rehabilitation programs focusing on psychological and behavioral aspects. In future, the cutoff value for high and low levels of pain behavior should be investigated in longitudinal studies.

### *Limitations:*

- The majority of cutoff values for pain behavior were based on literature to ensure their validity. However, our searches found no cutoff levels for some items. In these latter cases, cutoff levels were set at either the 10<sup>th</sup> or 5<sup>th</sup> percentile. Using these very conservative estimates meant that the number of patients with high levels of pain behavior would not be overestimated.

However, most of the items with a questionable cutoff were excluded during the RASCH analysis, due to their statistical performance, demonstrating that there were more problems inherent in these items. On the other hand the validity of cutoff values of the remaining PBA items was confirmed.

- In our study the PSI was 0.83, which justifies the application of the PBA both for groups and as acceptable for individuals. Commonly, a simple questionnaire completed by patients reaches a higher PSI value. However, assessments that are more complex and performed on the basis of various observations, such as the PBA, usually reach lower reliability scores [45]. In video-based ratings, reliability scores of 0.6 are acceptable for use in individuals.
- At this stage of research, the level of pain behavior can only be estimated by consulting the distribution of the total score. In comparison to the usual patients with chronic LBP, the patients in this study have a limited or impaired working capacity. This is the reason that they are specifically referred for a fitness-for-work evaluation. A case mix with more severely affected patients as in our study may lead to a higher prevalence of high levels of pain behavior than in usual samples with chronic LBP. Therefore, in this more affected sample it may be justified to judge scores above the 75<sup>th</sup> percentile as high

levels of pain behavior. The 75<sup>th</sup> percentile would correspond with 13 points on the total score.

## **Conclusion**

The PBA is a valid method to assess pain behavior in patients with CNSLBP and is valid for use in individuals. It may help to detect patients with high levels of pain behavior, improve the screening process for physically demanding rehabilitation programs and to allocate patients effectively to the right treatment.

Future research should investigate the inter-tester reliability of the PBA. Further longitudinal studies are needed to determine PBA cutoff values for high, medium and low risk of poor outcome during rehabilitation or for return to work. Prediction models would indicate whether different weighting of the subscales would improve prediction. Finally we suggest evaluation of the construct validity of the PBA in other settings and countries.

## **Acknowledgements**

This study was financially supported by the Swiss Association of Rehabilitation. We thank Karen Linwood for proofreading the manuscript.

## **Conflict of interest**

Authors, Katharina Meyer, Andreas Klipstein, Peter Oesch, Beatrice Jansen, Jan Kool and Karin Niedermann declare that they have no conflict of interest.

Ethical approval was obtained from the 3 regional ethics committees (EKSG 08/029/2B; EKZH SPUK N°784, EKAG 08/058) where the rehabilitation clinics are



located. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5). Informed consent was obtained from all patients prior to inclusion in the study.

## References

1. Pillastrini P, Gardenghi I, Bonetti F, Capra F, Guccione A, Mugnai R, Violante F. An updated overview of clinical guidelines for chronic low back pain management in primary care. *Joint Bone Spine* 2012;79(2):176-85.
2. Petit A, Roche-Leboucher G, Bontoux L, Dubus V, Ronzi Y, Roquelaure Y, Richard I. Effectiveness of three treatment strategies on occupational limitations and quality of life for patients with non-specific chronic low back pain: Is a multidisciplinary approach the key feature to success: study protocol for a randomized controlled trial. *BMC musculoskeletal disorders* 2014;15(1):131.
3. Hong J, Reed C, Novick D, Happich M. Costs associated with treatment of chronic low back pain: an analysis of the UK General Practice Research Database. *Spine* 2013;38(1):75-82.
4. Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? *JAMA*. 2010;303(13):1295-302.
5. van Middelkoop M, Rubinstein SM, Kuijpers T, Verhagen AP, Ostelo R, Koes BW, van Tulder M. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. *Eur Spine J*. 2011;20(1):19-39.
6. Royal Dutch Society for Physical Therapy. KNGF Guideline Low back pain. Amersfoort 2013(5); ISSN 1567-6137.
7. Sullivan MJ, Thorn B, Haythornthwaite JA, Keefe F, Martin M, Bradley LA, Lefebvre, J. Theoretical perspectives on the relation between catastrophizing and pain. *The Clinical journal of pain*. 2001;17(1):52-64.
8. Keefe FJ, Wilkins RH, Cook WA. Direct observation of pain behavior in low back pain patients during physical examination. *Pain* 1984;20(1):59-68.
9. Robertson L, Brodowicz G, Swafford A. Improved detection of submaximum effort in upper extremity strength and strength-endurance performance testing. *J Occup Rehab*. 1997;7(2):83-95.
10. Lemstra M, Olszynski W, Enright W. The sensitivity and specificity of functional capacity evaluations in determining maximal effort: a randomized trial. *Spine*. 2004;29(9):953-9.
11. BÄK, KBV, AWMF. National Disease Management Guideline 'Low back pain' – Short Version. 2011 Last amended: August 2013. Available from: <http://www.kreuzschmerz.versorgungsleitlinien.de>. Accessed Dec 2014.

12. Hampel P, Gemp S, Mohr B, Schulze J, Tlach L. Long-Term Effects of a Cognitive-Behavioral Intervention on Pain Coping among Inpatient Orthopedic Rehabilitation of Chronic Low Back Pain and Depressive Symptoms. *Psychother Psych Med*. 2014 (64):439-47.
13. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: Efficacy, innovations, and directions for research. *Amer Psychologist*. 2014;69(2):153.
14. Innes E. Handgrip strength testing: A review of the literature. *Aust Occup Ther J*. 1999;46(3):120-140.
15. Joughin K, Gulati P, Mackinnon SE, McCabe S, Murray JF, Griffiths S, Richards R. An evaluation of rapid exchange and simultaneous grip tests. *J Hand Surgery*. 1993;18(2):245-52.
16. Matheson L, Bohr P, Hart D. Use of maximum voluntary effort grip strength testing to identify symptom magnification syndrome in persons with low back pain. *Journal of Back and Musculoskeletal Rehabilitation*. 1998;10(3):125-35.
17. Oesch P, Meyer K, Jansen B, Mowinckel P, Bachmann S, Hagen KB. What is the role of "nonorganic somatic components" in functional capacity evaluations in patients with chronic nonspecific low back pain undergoing fitness for work evaluation? *Spine* 2012;37(4):E243-50.
18. Matheson LN, Isernhagen SJ, Hart DL. Relationships Among Lifting Ability, Grip Force, and Return to Work. *Physical Therapy*. 2002;82(3):249-56.
19. Kool J, Oesch P, de Bie R. Predictive tests for non-return to work in patients with chronic low back pain. *European Spine J*. 2002;11(3):258-66.
20. Oesch PR, Hilfiker R, Kool JP, Bachmann S, Hagen KB. Perceived functional ability assessed with the spinal function sort: is it valid for European rehabilitation settings in patients with non-specific non-acute low back pain? *Eur Spine J*. 2010;19(9):1527-33.
21. Waddell G, McCulloch JA, Kummel ED, Venner RM. Nonorganic physical signs in low-back pain. *spine*. 1980;5(2):117-25
22. Fishbain DA, Cole B, Cutler RB, Lewis J, Rosomoff HL, Rosomoff RS. A Structured Evidence-Based Review on the Meaning of Nonorganic Physical Signs: Waddell Signs. *Pain Medicine* 2003;4(2):141-81.
23. Matheson L. Symptom magnification syndrome structured interview: rationale and procedure. *J Occup Rehab*. 1991;1(1):43-56.
24. Schmidt CO, Lindena G, Pflingsten M, Kohlmann T, Chenot J-F. Vergleich zweier Screening-Fragebogen für Patienten mit Rückenschmerzen. *Der Schmerz* 2014;28(4):365-73.

25. Hart DL, Isernhagen SJ, Matheson LN. Guidelines for functional capacity evaluation of people with medical conditions. *J Orthop Sports Phys Ther.* 1993;18(6):682-6.
26. Innes E. Reliability and validity of functional capacity evaluations: an update. *International Journal of Disability Management.* 2006;1(01):135-48.
27. Matheson LN. Symptom Magnification Syndrom. In: Isernhagen SJ (ed). *Work Injury: management and prevention.* Gaithersburg: Aspen; 1988. p. 257-82.
28. Fordyce WE. Back pain in the workplace: management of disability in nonspecific conditions : a report of the Task Force on Pain in the Workplace of the International Association for the Study of Pain. International Association for the Study of Pain (ed), Seattle. IASP Press; 1995.
29. Abenham L, Rossignol M, Valat JP, Nordin M, Avouac B, Blotman F, Charlot J, Dreiser RL, Legrand E, Rozenberg S, Vautravers P. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine* 2000;25(4 Suppl):1s-33s.
30. Trippolini MA, Dijkstra PU, Jansen B, Oesch P, Geertzen JH, Reneman MF. Reliability of clinician rated physical effort determination during functional capacity evaluation in patients with chronic musculoskeletal pain. *J Occup Rehabil.* 2014;24(2):361-9.
31. Sindhu BS, King PM. Assessing Evaluatee Effort. In: Genovese E, Galper JS, eds. *Guide to the Evaluation of Functional Ability. How to Request, Interpret, and Apply Functional Capacity Evaluations:* Chicago, AMA American Medical Association; 2009. p. 195-226.
32. Shechtman O. Using the coefficient of variation to detect sincerity of effort of grip strength: a literature review. *J Hand Therapy* 2000;13(1):25-32.
33. Shechtman O, Anton SD, Kanasky Jr WF, Robinson ME. The use of the coefficient of variation in detecting sincerity of effort: a meta-analysis. *Work: A Journal of Prevention, Assessment and Rehabilitation.* 2006;26(4):335-41.
34. Gibson L, Strong J. The reliability and validity of a measure of perceived functional capacity for work in chronic back pain. *J Occup Rehab.* 1996;6(3):159-75.
35. Apeldoorn AT, Bosselaar H, Blom-Luberti T, Twisk JW, Lankhorst GJ. The reliability of nonorganic sign-testing and the Waddell score in patients with chronic low back pain. *Spine* 2008;33(7):821-6.

36. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*. 2007;57(8):1358-62.
37. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Brit J Clinical Psych*. 2007;46(1):1-18.
38. MacKnight C, Rockwood K. Rasch analysis of the hierarchical assessment of balance and mobility (HABAM). *Journal of clinical epidemiology*. 2000;53(12):1242-7.
39. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work ability index. 2nd ed. Helsinki; Finland: Finnish Institute of Occupational Health.1998.
40. Ilmarinen J, Tuomi K, Klockars M. Changes in the work ability of active employees over an 11-year period. *Scand J Work Environ Health*. 1997;23 Suppl 1:49-57.
41. Keefe FJ, Lefebvre JC, Egert JR, Affleck G, Sullivan MJ, Caldwell DS. The relationship of gender to pain, pain behavior, and disability in osteoarthritis patients: the role of catastrophizing. *Pain* 2000;87(3):325-34.
42. Andersson GBJ. Epidemiological features of chronic low-back pain. *The Lancet* 1999;354(9178):581-5.
43. Al-Obaidi SM, Beattie P, Al-Zoabi B, Al-Wekeel S. The relationship of anticipated pain and fear avoidance beliefs to outcome in patients with chronic low back pain who are not receiving workers' compensation. *Spine*. 2005;30(9):1051-7.
44. Banerji M, Smith RM, Dedrick RF. Dimensionality of an early childhood scale using rasch analysis and confirmatory factor analysis. *Journal of outcome measurement*. 1997;1(1):56-85.
45. Luomajoki H, Kool J, de Bruin ED, Airaksinen O. Reliability of movement control tests in the lumbar spine. *BMC musculoskeletal disorders*. 2007;8(1):90.